



Jay, T. (2012). First- and second-order reactivity to verbal protocols: An example from a study on strategy variability. *International Journal of Research & Method in Education*, 35(2), 181-194.
<https://doi.org/10.1080/1743727X.2012.689976>

Early version, also known as pre-print

Link to published version (if available):
[10.1080/1743727X.2012.689976](https://doi.org/10.1080/1743727X.2012.689976)

[Link to publication record in Explore Bristol Research](#)
PDF-document

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

This article was downloaded by: [University of Bristol]

On: 11 October 2012, At: 07:56

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



International Journal of Research & Method in Education

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/cwse20>

First- and second-order reactivity to verbal protocols: an example from a study on strategy variability

Tim Jay ^a

^a Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol, BS8 1JA, UK

Version of record first published: 24 May 2012.

To cite this article: Tim Jay (2012): First- and second-order reactivity to verbal protocols: an example from a study on strategy variability, International Journal of Research & Method in Education, 35:2, 181-194

To link to this article: <http://dx.doi.org/10.1080/1743727X.2012.689976>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.tandfonline.com/page/terms-and-conditions>

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

First- and second-order reactivity to verbal protocols: an example from a study on strategy variability

Tim Jay*

Graduate School of Education, University of Bristol, 35 Berkeley Square, Bristol BS8 1JA, UK

(Received 17 May 2011; final version received 5 February 2012)

Verbal reports are a common method of data collection in studies of mathematics learning, often in studies with a longitudinal component or those employing microgenetic methods where several observations of problem-solving are made over a short period of time. Whilst there is a fairly substantial literature on reactivity to verbal reports, addressing concerns that the production of verbal reports will change the behaviour that participants are reporting on, little of this literature addresses the question of how verbal reports might affect *subsequent* trials (what we would like to call *second-order* reactivity). In order to address the issues of both first- and second-order reactivity, ninety-nine 14-year-old participants were asked to provide answers to a set of six questions on two occasions, one week apart. Fifty-eight participants were asked for written answers and explanations, while 41 participants were asked for both concurrent and retrospective verbal reports. The results show evidence for both first- and second-order reactivity. The distribution of strategies employed in the first session was different for the two conditions and there were some interesting differences in patterns of strategy selection across the two sessions. These findings have some important implications for the use of verbal reports over multiple trials.

Keywords: verbal reports; methodology; strategy

Introduction

The use of verbal reports in educational research

In order for researchers to determine as accurately as possible the strategies used by children and adults to solve problems, verbal reports are often used. Verbal reports have been a core component of the data in studies in a variety of domains, including comprehension of native (Wade 1990) and second-language (Block 1986) texts, in scientific reasoning (Klahr, Fay, and Dunbar 1993), memory (McGilly and Siegler 1989), and arithmetic in both children (Siegler 1987) and adults (LeFevre, Sadesky, and Bisanz 1996). Verbal reports have also been used to investigate thinking processes in less traditional domains including entrepreneurship (Sandberg, Schweiger, and Hofer 1988) and library catalogue searching (Morrison 1999). This paper focuses on the use of verbal reports to study strategy choice in response to mathematical problems, especially over multiple trials – examples of this in the literature can be found in a

*Email: tim.jay@bristol.ac.uk

number of papers (Katz, Bennett, and Berger 2005; Luwel, Siegler, and Verschaffel 2008).

Verbal reports can take different forms. Verbal data can be collected concurrently with problem-solving, either by prompting participants with questions or by asking participants to think aloud the whole time that they are working on a problem (Ericsson and Simon 1993; van Someren, Barnard, and Sandberg 1994). Alternatively, data can be collected after children have completed a problem, via retrospective prompting. There is evidence to suggest that a combination of think-aloud protocols and retrospective verbal reports is the best means for gathering data regarding children's uses of strategy when solving mathematical problems (Kuusela and Paul 2000). The main advantage of collecting data while children are actually solving problems is to prevent the loss of information through children's memory failure when reporting problem-solving procedures at the end of an activity (Wade 1990). The main disadvantage of collecting verbal data while children are working on problems is that it poses a risk of disrupting or altering the problem-solving process that is being measured. Think-aloud protocols as described by Ericsson and Simon (1993) seem to address both of these issues – they are claimed to provide data on children's cognitive processes while they are occurring without disrupting or altering those processes.

Kuusela and Paul (2000) compared the effectiveness of think-aloud protocols and retrospective reports. They found that concurrent think-alouds outperformed retrospective reports in general, eliciting more statements and better insight into problem-solving processes. However, they concluded that the use of retrospective reports could be a valuable *addition* to think-alouds. Retrospective reports provided more information about children's final decisions in the problem-solving process and could provide useful support for the data elicited by the concurrent think-aloud protocols.

Threats to validity

There are some limitations of think-aloud protocols that should be borne in mind during their use. One major criticism of these data can be found in Nisbett and Wilson (1977), where it is claimed that participants do not necessarily have direct access to the cognitive processes that they are being asked to report on. Verbal reports of those processes will only be accurate where 'influential stimuli are salient and are plausible causes of the responses they produce' (231). Some aspects of problem-solving processes may be automatic, or otherwise happen too quickly for participants to be sufficiently aware of their nature to be able to report on them, for example.

Young (2005) lists three methodological issues involved in the use of think-alouds; reactivity, participants' verbal ability, and validity in analysis. Participants' verbal abilities can cause a problem due to their variation within a sample. Children can vary widely in their ability to provide a verbal report on their cognitive processes. Wilson (1994) and Branch (2000) both describe research situations in which the variation of the completeness of protocols has caused difficulties in interpreting the data. A significant threat to validity in the use of think-aloud protocols is the fact that children are not necessarily able to verbalize the entire cognitive process that they experience when solving a problem. We know, for example, that there are unconscious aspects of strategy discovery and children can begin to use a strategy some time before they have the ability to verbalize an understanding of it (Siegler and Stern 1998). Also, there are aspects of problem-solving that are automatic and therefore unlikely to be reported (Wade 1990).

The main issue of validity with respect to this article is reactivity. The problem of reactivity is that the use of think-alouds might have the effect of altering the problem-solving processes that those protocols are intended to inform on. Stratman and Hamp-Lyons (1994) discuss a few causes of reactivity when using think-aloud protocols. For example, when participants in a study are asked to think aloud, they are not working on a problem in their usual way. They are being asked to complete an additional task, which may have an impact on their available cognitive resources. Participants are also having their attention drawn to the cognitive processes involved in the task set, to a degree which they would probably not in normal circumstances.

Second-order reactivity

One question that has not been addressed so much in the literature is whether there is reactivity to verbal reports outside of a single trial. In this paper, we make the distinction between 'first-order reactivity', where behaviour is affected by the use of verbal reports within a single trial, and 'second-order reactivity', where the use of verbal reports in a trial affects behaviour in subsequent trials. This is an important issue, as verbal reports are frequently used in research involving multiple sessions, such as in longitudinal research or research using microgenetic methods. If the use of verbal reports in a particular trial causes changes to behaviour on subsequent trials, this could have profound implications for the interpretation of findings in any study that uses verbal reports on multiple trials.

There is some evidence in the literature that suggests that we could expect to see second-order reactivity to verbal protocols. In this section, we focus on two aspects of this literature, verbal overshadowing and self-explanation effects. 'Verbal overshadowing' is the name given to the phenomenon whereby people asked to give a verbal description of a face, for example, later find it more difficult to recognize that face. The memory of the verbal description produced at the time of seeing the face *overshadows* the memory of the face itself (Dodson, Johnson, and Schooler 1997). The verbal report produced at the time of seeing the face does not affect behaviour at that time, but rather affects behaviour on a later occasion. It is possible that a verbal overshadowing effect could play a part in problem-solving activity. By analogy with verbal overshadowing of memory for faces, it may be that if participants are asked to produce a verbal report of their problem-solving, it is possible that memory for that verbal report then overshadows memories of the problem or of the problem-solving activity itself. This kind of effect would be observable only across trials, and not within a single trial, as it depends on an effect of memory rather than one of perception.

Literature on self-explanation effects also suggests that we might expect to see second-order reactivity to verbal reports. When participants are prompted to explain material to themselves after reading it, they are more likely to recall that material at a later date (Chi et al. 1994). Similar to the verbal overshadowing effect, self-explanation effects seem to suggest that verbal reports could have implications for the validity of verbal reports across trials. If verbal reports are used in order to inform the researcher about problem-solving processes and strategy choices, then it is important that we understand whether and how those reports might influence the processes and choices that are being researched.

The present study

It is clear that confidence in the validity of verbal protocols is essential for their continued use in research in cognition. As part of a larger body of research (Jay 2007), an

experiment was conducted in which children were presented with two sets of problems, one week apart. The problems involved children making judgments about relative rates of change, on the basis of straight-line graphs.

Participants in the experiment differed only in terms of the absence or presence of verbal reports. In one condition, children were asked to write down their answer to each problem, and also an explanation for their answer. In the second condition, children were asked to think aloud while working on each problem. They were also asked to give a retrospective spoken report on the set of problems once they were completed.

The focus on analysing patterns of strategy use across sessions offered an ideal opportunity to examine any effects of verbal protocols on children's thinking. Not only was it possible to analyse the effect of verbal protocols on concurrent behaviour, but it was also possible to identify any effect on subsequent problem-solving behaviour. In a great deal of research, especially that research using microgenetic methods, verbal reports are employed in longitudinal studies in order to track the development of children's response to problems from a particular domain. If verbal reports have any effect on children's problem-solving behaviour across sessions, it is vital that we monitor and understand that effect.

Method

Design

Participants were asked to solve a set of six problems in each of two sessions, one week apart. These problems involved reasoning on the basis of straight-line graphs (an example of the problems can be found in Appendix A). No feedback was given to participants regarding their performance. The experiment employed a between-groups design with two conditions of the independent variable, which concerned the conditions under which participants provided answers to the problem sets; one group of participants were asked to give their answers and explanations in writing whilst the second group were asked to think aloud, giving answers and explanations verbally. The second group was also asked for retrospective verbal reports, following the solution of each problem. The dependent variable was the set of strategies used in each session – this was used in order to assess first-order reactivity. If there is no reactivity to verbal protocols, then we would expect a similar distribution of strategies to be used under both conditions, whereas reactivity would lead to a different distribution of strategies between the two conditions. A second dependent variable was derived from the difference between the sets of strategies reported during the first session and during the second session – assessing second-order reactivity. This was achieved in two ways – by calculating the correlation between uses of particular strategies across sessions and by comparing distribution of patterns of strategy use across sessions (see Results section for further descriptions of these analyses).

Participants

Participants were ninety-nine 13–14-year-old children. 58 children (24 girls and 34 boys) were in the written-answer group, while 41 children (16 girls and 25 boys) were in the think-aloud group. Participants were drawn from four complete classroom groups, two each from two local schools. Two complete classes (one from each school) took part in each condition. All of the children were expected to achieve level 5 in the National Curriculum Mathematics Test (SAT), the standard mathematics test taken by

14-year-olds in schools in England. Possible grades in this Year 9 SAT range from level 3 to level 8, and level 5 represents an approximately average level of performance. Pilot studies had shown that children at this level of attainment had access to the problem sets (children understood the questions and could come up with reasonable explanations for their answers) but had not yet developed stable strategies for finding answers. The children had not encountered rate of change problems in the classroom before.

Whilst participants within each of the two conditions were matched according to classroom mathematics performance and to gender, it is possible that there may have been other aspects of participants' mathematical behaviours and abilities that may have systematically varied between classes. However, given the nature of the tasks involved in this study, and in particular the fact that this type of task had not been encountered by any participants in their mathematics lessons before the study took place, it was considered that there was a minimal risk of difference amongst groups. Given that there was still a risk, no matter how small, analyses of the findings presented below include tests conducted in order to assess the comparability of groups from different schools. This is based on the argument that differences between schools are likely to be larger than between classes within a school.

Tasks

Two isomorphic sets of problems were used (see Appendix A), adapted from an experiment reported in Mevarech and Stern (1997). All problems involved children making judgements about rates of change on the basis of linear graphs.

There were six questions, three each associated with two different graphs, taken directly from Mevarech and Stern (1997) with only the wording changed in order to improve children's understanding of the questions (based on previous pilot work). Each graph showed two straight lines that crossed at a point. These asked children about the rate of change of the two lines on the graph and also asked children to explain how they decided on their answer. For example, one question asked children, 'after 1984, did the income of company A increase faster, slower or at the same rate as the income of company B'? Another question asked, 'in 1984, was there a change in the rate of increase of income of company A'?

The instructions advised children that they could write on the graphs if they thought it might help, and also that they should pay careful attention to explanations that were asked for.

The only difference between the two sets of problems was the context. For example, the sparse context problems involved a graph with axes labelled 'x' and 'y' and lines labelled 'line A' and 'line B', while the realistic context problems involved a graph with axes labelled 'income' and 'year' and lines labelled 'company A' and 'company B'.

Procedure

Under both conditions, children were tested individually in a quiet space outside of their usual classroom. This was done in order to ensure that any observed difference between the groups could not be argued to be due to the difference between working on the problems individually and in a group in a classroom. Children were administered two sets of problems, one week apart. The two sets of problems were isomorphic, differing only in terms of the context attached to the problems. One half of the participants

were administered the two sets of problems in the reverse order to the other half of participants. The problem sets took approximately 15 min to complete in total.

Written explanations condition

Participants were asked to complete the set of problems, writing their answer in the box provided, along with an explanation for their answer. Once the set of problems was completed, the experimenter checked the explanations given by the children to ensure that enough detail had been given in order to determine the strategy employed.

The experimenter had the opportunity to ask each child for more information at the end of each session in order to elicit missing answers and to obtain further information regarding strategy. Prompts were given with the intention that children should not be led towards one explanation or another and were as neutral as possible, such as 'please make sure you give an explanation for each question' or 'could you write a bit more about how you did this one'. The experimenter's responses were also as neutral as possible, intended to give no indication as to the correctness or otherwise of answers or of explanations. In practice, these interventions were only required for three participants.

Verbal reports condition

Under this condition, children were asked to think aloud whilst working on each problem. They were also asked for a retrospective report for each problem. The usage of verbal protocols in this experiment followed the instructions given in Ericsson and Simon (1993). Before children saw any problems, they were given some training in think-aloud protocols. This also took the form described by Ericsson and Simon (1993), whereby participants were given three problems to work on, while thinking aloud:

- Add together the numbers 26 and 48.
- How many windows are there in your house?
- Name 20 different animals.

If children were not thinking aloud appropriately during any of these problems, the experimenter gave an example of how thinking aloud for that problem could sound. All children were thinking aloud appropriately, according to the criteria in Ericsson and Simon (1993), for the practice problems by the end of the training session.

Participants were asked to think aloud whilst working on each of the six problems in the set. The experimenter did not interrupt at any point, except to remind the participant to keep talking if there was a period of more than 15 s of silence. On completion of each problem, the participant was asked to give a retrospective verbal report, explaining how they had arrived at their solution strategy. Sessions for this condition were recorded using a video camera, positioned so that it gave a view over the participant's right shoulder. The video recording was used in order to code the strategies reported by participants.

Results

The purpose of this analysis is to examine ways in which the use of verbal reports may be associated with changes in strategy selection by participants. There are two aspects to this analysis. The first is to look at the distribution of strategies selected on the first trial, under both conditions. Given that the majority of the literature on verbal protocols

has shown reactivity to be a rare phenomenon, we should not expect to see substantial differences between conditions. The second aspect of this analysis involves looking at the way in which the set of strategies selected in the second trial differs to the set selected in the first trial (second-order reactivity). This is an aspect of reactivity that has generally not been explored in the literature, but one where we might expect to see substantial differences between the two conditions.

Comparability of samples

Here, the strategies reported by participants from different schools will be compared. The purpose of this analysis is to generate confidence in the assertion that the samples taking part in each condition of the experiment (written-answers and verbal-reports conditions) are comparable. As discussed above, this is important in order to be able to ascribe any observed differences to the experimental condition rather than individual differences. The argument will be that if samples taken from different schools are comparable then it is highly likely that samples under the two conditions are comparable.

In order to assess the comparability of samples, the distribution of strategies reported in the first session for participants from school A and school B were compared, using a chi-squared test. This was done for both the written-answers group and the verbal-reports group. For the written-answers group, there was no significant difference between the 2 schools ($\chi^2 = 1.49$, $df = 5$, $p > 0.05$). For the verbal-reports group, there was also no significant difference between the 2 schools ($\chi^2 = 5.89$, $df = 4$, $p > 0.05$). We can therefore conclude that the samples from different schools within each condition are comparable with one another.

Distribution of strategies used in trial 1: First-order reactivity

The strategies used by children in the written-answers condition, with their distribution across instances of strategy use, are shown in Figure 1. The equivalent data for the verbal-reports condition are given in Figure 2. Children in both conditions were most likely to use explanations involving the relative height or steepness of the lines in order to solve the problems. Visual inspection of the two figures reveals a strong similarity between the two distributions. There is a similar effect of gender under both conditions, whereby boys were more likely to use a strategy based on the steepness of the lines, and girls were more likely to use a strategy based on their height.

The main difference between the two distributions consists in the fact that a greater proportion of participants selected one of the two most popular strategies in the verbal-reports condition than in the written-answers condition (54% in the written-answers condition, 80% in the verbal-reports condition). The difference in distribution of strategies used between the two conditions (the difference between the distribution shown in Figure 1 and that shown in Figure 2) was tested using a chi-squared test,¹ and despite the visual similarity was shown to be significant ($\chi^2 = 57.2$, $df = 4$, $p < 0.01$).

Comparison of strategies used across trials: second-order reactivity

This analysis made use of a measure of strategy variability across trials. To assess the level of variability of strategy use of boys and girls, correlations were taken between children's use of particular strategies in the first and second sessions. This analysis was conducted for each of the two most commonly used strategies, the use of relative

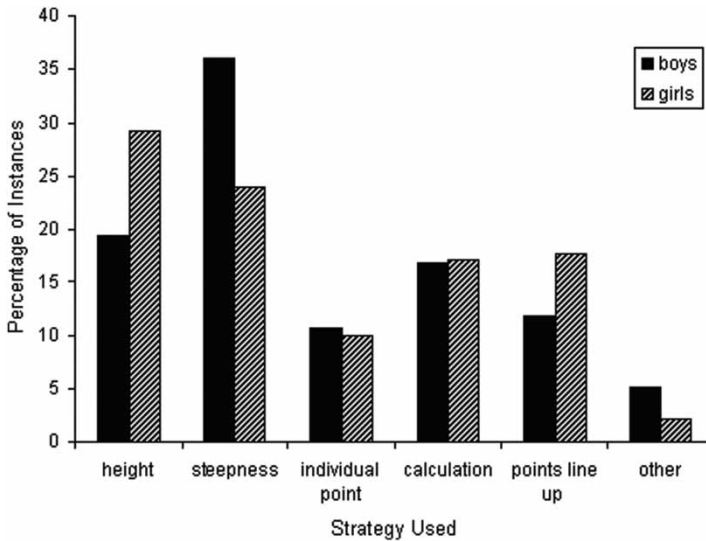


Figure 1. Distribution of strategies used in trial 1 – written-answers condition.

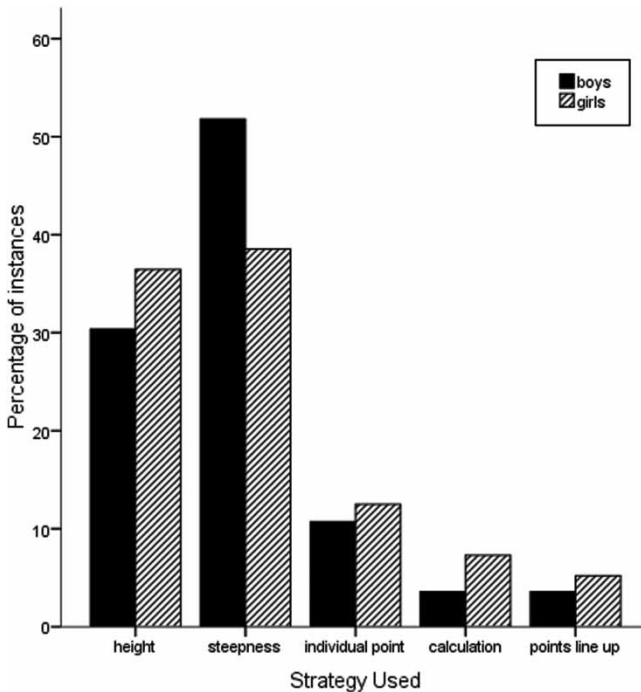


Figure 2. Distribution of strategies used in trial 1 – verbal-reports condition.

height and the use of relative steepness. So, for example, to assess variability with respect to steepness, the correlation between the number of times (out of 6) that a strategy involving steepness was used in the first and second sessions was calculated. Therefore, a high correlation would indicate low variability (steepness is used on a similar

Table 1. Pearson correlation coefficients of uses of strategies involving height and steepness by gender and condition.

	Written answers		Verbal reports	
	Height	Gradient	Height	Gradient
Boys	0.059	0.180	0.713**	0.623**
Girls	0.694**	0.504*	0.120	0.176

*Indicates significant at $p = 0.05$.

**Indicates significant at $p = 0.001$.

number of occasions in both sessions) and a low correlation would indicate high variability. For strategies other than those involving steepness or height, there were too few instances of use in order to generate a reliable correlational analysis. These analyses are summarized in Table 1.

The correlations presented in Table 1 suggest that there is a difference in the ways that boys and girls responded to the two conditions. In the written-answers condition, boys showed a higher level of strategy variability than did girls (higher correlation indicates lower strategy variability). In the verbal-reports condition, girls showed a higher level of strategy variability than did boys.

Fisher's z transformations were used to test the differences in correlations, between conditions and between genders. In the written-answers condition, the difference between girls and boys was significant both for strategies involving steepness ($z = 1.32$, $p < 0.05$) and strategies involving height ($z = 2.82$, $p < 0.05$). In the verbal-reports condition, the difference between boys and girls was again significant (for steepness, $z = 1.48$, $p < 0.05$; for height, $z = 2.08$, $p < 0.05$). Across conditions, there were also some significant differences: for the boys, there are significant differences in correlations of both strategies involving height ($z = 2.52$, $p < 0.05$) and strategies involving steepness ($z = 1.68$, $p < 0.05$), while for girls, there is a significant difference in correlation of strategies involving height ($z = 2.41$, $p < 0.05$) and a marginally significant difference (significant at $p = 0.05$ if using a 1-tailed test) in correlation of strategies involving steepness ($z = 1.24$, $p < 0.1$).

Supplementary data – test scores

Additional data were analysed in order to clarify some details of these findings. The average number of correct answers given in each session, under each condition, is given in Table 2.

A $2 \times 2 \times 2$ ANOVA (gender \times condition \times session, with repeated measures for session) was conducted, with the number of correct answers as dependent variable. The only significant main effect was that for gender, $F(1,95) = 9.17$, $p = 0.003$. All other main effects and all interaction effects were not significant. Crucially, there was no significant main effect of condition, $F(1,95) = 1.02$, $p = 0.31$.

Results summary

There was evidence of first-order reactivity in the data. The distribution of strategies employed was similar in session 1 under both conditions, but participants in the verbal-reports condition were more likely to use one of the two most common

Table 2. Average scores (out of 6), with standard deviations in parentheses.

	Written answers		Verbal reports	
	Session 1	Session 2	Session 1	Session 2
Boys	4.57 (1.50)	5.00 (1.23)	4.31 (1.82)	4.56 (1.41)
Girls	3.92 (1.28)	3.95 (1.55)	3.76 (1.48)	3.64 (1.63)

strategies, using either the relative heights or relative gradients of the lines. In addition to this first-order reactivity, however, there was also evidence of second-order reactivity – where the use of verbal reports is associated with changes in behaviour in a later session. In the written-answers condition, boys were more likely to choose a new set of strategies in session 2, while girls were more likely to use the same set of strategies in session 2 as they did in session 1. In the verbal-reports condition, the opposite effect was observed, whereby boys were more likely to use the same strategies in session 2 as used in session 1 and girls were more likely to select a new set of strategies.

Discussion

A comparison of the two conditions in this experiment provides evidence for both first- and second-order reactivity to verbal protocols. The only procedural difference between the two conditions was the use of verbal protocols in one of them. In both conditions, participants were at the same level of achievement (level 5 on Year 9 mathematics SAT), had the same experience of the problem set used, and were asked to complete the problem set individually. Identical problem sets were used in each study.

The difference between the two conditions is striking for a number of reasons, not least because of a widespread assumption in research on learning that reactivity does not occur. There is a fairly substantial literature that suggests there should be no reactivity to think-aloud protocols (Ericsson and Simon 1993; Kirk and Ashcraft 2001; Kuusela and Paul 2000; Robinson 2001; Wilson 1994). The patterns of strategy variability observed in the written-answers condition was the reverse of those observed in the verbal-reports condition. Not only was there a significant difference in strategy variability between boys and girls in both conditions, but there was a significant difference in strategy variability between boys in each of the two conditions and between girls in each of the two conditions. It seems that the only explanation for these findings is that children's problem-solving behaviour in the second session changed due to the use of verbal protocols in one of the two experimental conditions. It is important to note here that without the comparison of analyses of strategies across sessions, with and without verbal protocols, reactivity could not have been detected.

Possible mechanisms of reactivity

Mechanisms of first-order reactivity

The most important issue to be considered here are what qualities of this experimental situation might be implicated in the observed reactivity to verbal protocols. Stratman and Hamp-Lyons (1994) suggest that reactivity to verbal protocols can occur for two main reasons. When verbal protocols are employed, children have to complete an additional task, adding to the cognitive load. Also, they are having their attention drawn to the cognitive processes involved in the problem set.

The problems given to children in the pair of experiments discussed here subject participants to what could be considered a relatively high level of cognitive demand. Verbal protocols are often used to elicit information on strategy and procedure in comparatively simple domain, such as arithmetic (Siegler 1987). However, the list of domains in which verbal protocols have been used, given in the introduction to this article (including scientific reasoning, native- and second-language comprehension, and library catalogue searching) could certainly be considered comparable in terms of cognitive demand with those problems used here. In fact, it could be argued that any problem set for which there are competing strategies for solution involves a high level of cognitive demand. It is therefore possible that the increased cognitive load due to the use of verbal protocols caused children to behave differently in solving problems in this case. The participants' task, of solving an unfamiliar problem, and of giving a concurrent verbal report (also a relatively novel task for most participants), meets the definition of high intrinsic cognitive load as described in Paas, Renkl, and Sweller (2003). However, it is not clear why a high level of cognitive load would cause the differences in distribution of strategies observed in this study. In fact, as there was no difference in the number of questions answered correctly between the two conditions, and as increased cognitive load would normally be associated with a reduction in performance, it seems difficult to argue that increased cognitive load has contributed to the observed reactivity.

On the other hand, when verbal protocols are employed, children are also being made more aware of the cognitive processes involved in solving problems. The literature describing the effects of self-explanation on understanding (Chi et al. 1994) may be more useful than that on cognitive load in the explanation of the effects of verbal reports observed here. There is evidence to suggest that one result of self-explanation is that learners are more likely to reject certain strategies (Chi 2000; Williams and Lombrozo 2010) in a way that would lead to the change in distribution of strategies observed in the present study. Williams and Lombrozo (2010) asked participants to work on a task involving the categorization of alien robots. Participants studied robots that had been sorted into two categories. Robots were categorized by experimenters such that a more obvious rule (that Williams and Lombrozo call the '75% rule') could explain the category membership of 75% of robots, whilst a more subtle rule (that Williams and Lombrozo call the '100% rule') could explain the membership of 100% of robots. They found that participants asked to self-explain during their study were more likely to reject the more obvious 75% rule. This provides a useful analogy with the present study, where self-explanation could have been a means to encourage participants to reject some potential strategies – leading to the observed increase in the proportion of trials on which the two most common strategies, involving height or gradient, were used.

Mechanisms of second-order reactivity

This aspect of the findings is more difficult to interpret, as there is so little reference to second-order reactivity in the research methods literature. It is also possible that there are several interacting causes that are jointly responsible for the observed effects. As this reactivity is revealed through analysis by gender, it may be useful to consider known gender differences that may be implicated in the effect. Two that may be relevant relate to verbal abilities and to maths anxiety. Both of these possibilities will be discussed in this section.

Research has demonstrated that boys and girls differ in their affective response to mathematical problems. Girls tend to show more anxiety in response to mathematical

problems than do boys (Lussier 1996; Miller and Bichsel 2004). It has also been demonstrated that increased anxiety reduces performance by reducing available working memory capacity (Ashcraft and Kirk 2001; Miller and Bichsel 2004). It is more than possible that the use of verbal protocols increases the perceived pressure on a participant to perform well, thus raising levels of anxiety. We know that the occurrence and effects of anxiety are related to gender, so we might also expect to see differential reactions to verbal protocols according to gender. The findings in the present study indicate that, when asked to think aloud, girls were less likely to use the same set of strategies in the second session as they did in the first. This might be what we would expect to see, knowing that working memory capacity was reduced during the verbal-reports condition – as encoding of strategy use for long-term storage would be impaired by the reduction of working memory capacity. However, this explanation does not tell us why the reverse effect was observed for the boys.

It is well established in the literature that girls and boys differ in their verbal abilities. Girls tend to outperform boys. Particularly relevant to the present study is the finding of Hyde and Linn (1988), in a meta-analysis of 165 studies on gender differences in verbal abilities, that the largest effect size ($d = 0.33$) of a gender difference in verbal abilities was that for speech-production. It is not clear, however, exactly how this phenomenon might contribute to effect observed in the present study. Any advantage in verbalizing did not have an effect on girls' ability to answer the questions correctly, relative to the boys, in either the first or the second session.

Neither differences in verbal ability nor differences in maths anxiety offer an entirely satisfactory explanation for the observed effects. It is clear that further work is needed in order to understand the mechanism of second-order reactivity both in this specific study, and more generally. The importance of this study at this stage resides in the identification of an effect of second-order reactivity that would not have been identified without a control group or without analysis of group differences across sessions.

Implications of these findings

There are four main implications of the observed reactivity to think-aloud protocols. The first is that it shows that the gender effect on strategy variability, while fairly robust, appears to be somewhat dependent on the context in which problem-solving takes place and/or concurrent demands on children's cognitive resources. The most interesting aspect about this is that reactions to verbal protocols are very different for boys and girls. The employment of verbal protocols as a means of data collection appears to cause the strategy variability of girls to increase and of boys to decrease.

Focusing more on the issue of reactivity itself, there are three main sets of implications to be considered on the basis on the findings presented here. The first concerns the need to learn more about children's reactivity to verbal protocols, the second concerns the use of verbal protocols as a research method, and the third concerns the role of verbal protocols and think-aloud methods in teaching. It is clear that further work must be done in order to determine the nature and causes of any changes in behaviour due to verbal protocols.

A large amount of research is conducted that involves the analysis of changing strategy use over time. Some examples were given in the introduction to this article. Few experiments that employ verbal protocols involve the use of a control group for whom verbal protocols are *not* used. The findings presented here show that it can be

very difficult to identify reactivity without the use of such a control group. If either version of the experiment described in this article had been conducted independently of the other, the results would have been at best incomplete and at worst misleading.

We need to be as certain as possible that when we are using verbal protocols to explore cognition, we have a clear understanding of the processes that those protocols are describing. If the verbal protocols themselves are causing change in the processes we are interested in, they lose value as a research method. Unless it is possible to say for certain that verbal protocols are not causing change in the processes we are interested in, the use of a control group must be considered a necessity.

In relation to teaching, one implication of these findings is that children's talk about their own problem-solving activity can be seen to have had an effect of their future approach to similar problems. This can be interpreted alongside the self-explanation literature referred to earlier in this paper (Chi 2000; Chi et al. 1994; Williams and Lombrozo 2010), where we see that self-explanation can help learners to both reject strategies deficient in accuracy and/or efficiency and to generate alternative candidates.

Conclusion

The findings presented here demonstrate an example of reactivity to verbal protocols, differentiated according to gender. Further work must be done in order to establish the mechanism by which this effect occurs, but it seems more than possible that the increased demands on participants and the accompanying reduction in cognitive capacity play a part.

The use of verbal protocols in research in cognition must be given careful consideration. Without a control group of participants for whom verbal protocols are not used, the findings presented here demonstrate that an experiment could produce misleading results, especially where strategies across sessions are compared.

Note

1. Whilst Figures 1 and 2 show the distribution of strategies as percentages, the chi-squared test was conducted using frequencies.

References

- Ashcraft, M.H., and E.P. Kirk. 2001. The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology-General* 130, no. 2: 224–37.
- Block, E. 1986. The comprehension strategies of second language readers. *TESOL Quarterly* 20, no. 3: 463–94.
- Branch, J.L. 2000. The trouble with think alouds: Generating data using concurrent verbal protocols. Paper presented at the CAIS 2000: Dimensions of a global information science, University of Alberta.
- Chi, M.T.H. 2000. Self-explaining: The dual process of generating inference and repairing mental models. In *Advances in instruction psychology: Educational design and cognitive science*, Vol. 5, ed. R. Glaser. New Jersey, USA: Lawrence Erlbaum Associates.
- Chi, M.T.H., N. Deleeuw, M.H. Chiu, and C. Lavancher. 1994. Eliciting self-explanations improves understanding. *Cognitive Science* 18, no. 3: 439–77.
- Dodson, C., M. Johnson, and J. Schooler. 1997. The verbal overshadowing effect: Why descriptions impair face recognition. *Memory & Cognition* 25, no. 2: 129–39.
- Ericsson, K.A., and H.A. Simon. 1993. *Protocol analysis – verbal reports as data*. Cambridge, MA: MIT Press.
- Hyde, J., and M. Linn. 1988. Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin* 104, no. 1: 53–69.

- Jay, T. 2007. Explaining individual differences in strategy variability amongst secondary school mathematics students. Unpublished doctoral dissertation, University of Nottingham, Nottingham, UK.
- Katz, I., R. Bennett, and A. Berger. 2005. Effects of response format on difficulty of SAT-mathematics items: It's not the strategy. *Journal of Educational Measurement* 37, no. 1: 39–57.
- Kirk, E.P., and M.H. Ashcraft. 2001. Telling stories: The perils and promise of using verbal reports to study math strategies. *Journal of Experimental Psychology-Learning Memory and Cognition* 27, no. 1: 157–75.
- Klahr, D., A.L. Fay, and K. Dunbar. 1993. Heuristics for scientific experimentation: A developmental study. *Cognitive Psychology* 25: 111–46.
- Kuusela, H., and P. Paul. 2000. A comparison of concurrent and retrospective verbal protocol analysis. *American Journal of Psychology* 113, no. 3: 387–404.
- LeFevre, J., G.S. Sadesky, and J. Bisanz. 1996. Selection of procedures in mental addition: Reassessing the problem size effect in adults. *Journal of Experimental Psychology-Learning Memory and Cognition* 22, no. 1: 216–30.
- Lussier, G. 1996. Sex and mathematical background as predictors of anxiety and self-efficacy in mathematics. *Psychological Reports* 79, no. 3: 827–33.
- Luwel, K., R. Siegler, and L. Verschaffel. 2008. A microgenetic study of insightful problem solving. *Journal of Experimental Child Psychology* 99, no. 3: 210–32.
- McGilly, K., and R.S. Siegler. 1989. How children choose among serial recall strategies. *Child Development* 60: 172–82.
- Mevarech, Z., and E. Stern. 1997. Interaction between knowledge and contexts on understanding abstract mathematical concepts. *Journal of Experimental Child Psychology* 65, no. 1: 68–95.
- Miller, H., and J. Bichsel. 2004. Anxiety, working memory, gender, and math performance. *Personality and Individual Differences* 37, no. 3: 591–606.
- Morrison, H.G. 1999. Online catalogue research and the verbal protocol method. *Library Hi Tech* 17, no. 2: 197–206.
- Nisbett, R.E., and T.D. Wilson. 1977. Telling more than we can know: Verbal reports on mental processes. *Psychological Review* 84, no. 3: 231–59.
- Paas, F., A. Renkl, and J. Sweller. 2003. Cognitive load theory and instructional design: Recent developments. *Educational Psychologist* 38, no. 1: 1–4.
- Robinson, K.M. 2001. The validity of verbal reports in children's subtraction. *Journal of Educational Psychology* 93, no. 1: 211–22.
- Sandberg, W.R., D.M. Schweiger, and C.W. Hofer. 1988. The use of verbal protocols in determining venture capitalists' decision processes. *Entrepreneurship Theory and Practice*, 8–20.
- Siegler, R.S. 1987. The perils of averaging data over strategies – an example from children's addition. *Journal of Experimental Psychology-General* 116, no. 3: 250–64.
- Siegler, R.S., and E. Stern. 1998. Conscious and unconscious strategy discoveries: A microgenetic analysis. *Journal of Experimental Psychology-General* 127, no. 4: 377–97.
- Stratman, J.F., and L. Hamp-Lyons. 1994. Reactivity in concurrent think-aloud protocols: Issues for research. In *Speaking about writing: Reflections on research methodology*. ed. P. Smagorinsky, 89–111. Thousand Oaks, CA: Sage.
- van Someren, M.W., Y.F. Barnard, and J.A.C. Sandberg. 1994. *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.
- Wade, S.E. 1990. Using think alouds to assess comprehension. *The Reading Teacher* 43, no. 7: 442–51.
- Williams, J., and T. Lombrozo. 2010. The role of explanation in discovery and generalization: Evidence from category learning. *Cognitive Science* 34, no. 5: 776–806.
- Wilson, T.D. 1994. The proper protocol: Validity and completeness of verbal reports. *Psychological Science* 5, no. 5: 249–52.
- Young, K.A. 2005. Direct from the source: The value of 'think-aloud' data in understanding learning. *Journal of Educational Enquiry* 6, no. 1: 19–33.